SUPPLEMENTAL DATA

1. **SUPPLEMENTAL FIGURES**
2. **SUPPLEMENTAL FIGURE LEGENDS**
3. **SUPPLEMENTAL TABLES**
4. **SUPPLEMENTAL EXPERIMENTAL PROCEDURES**
5. **SUPPLEMENTAL REFERENCES**

SUPPLEMENTAL FIGURES

Figure S1

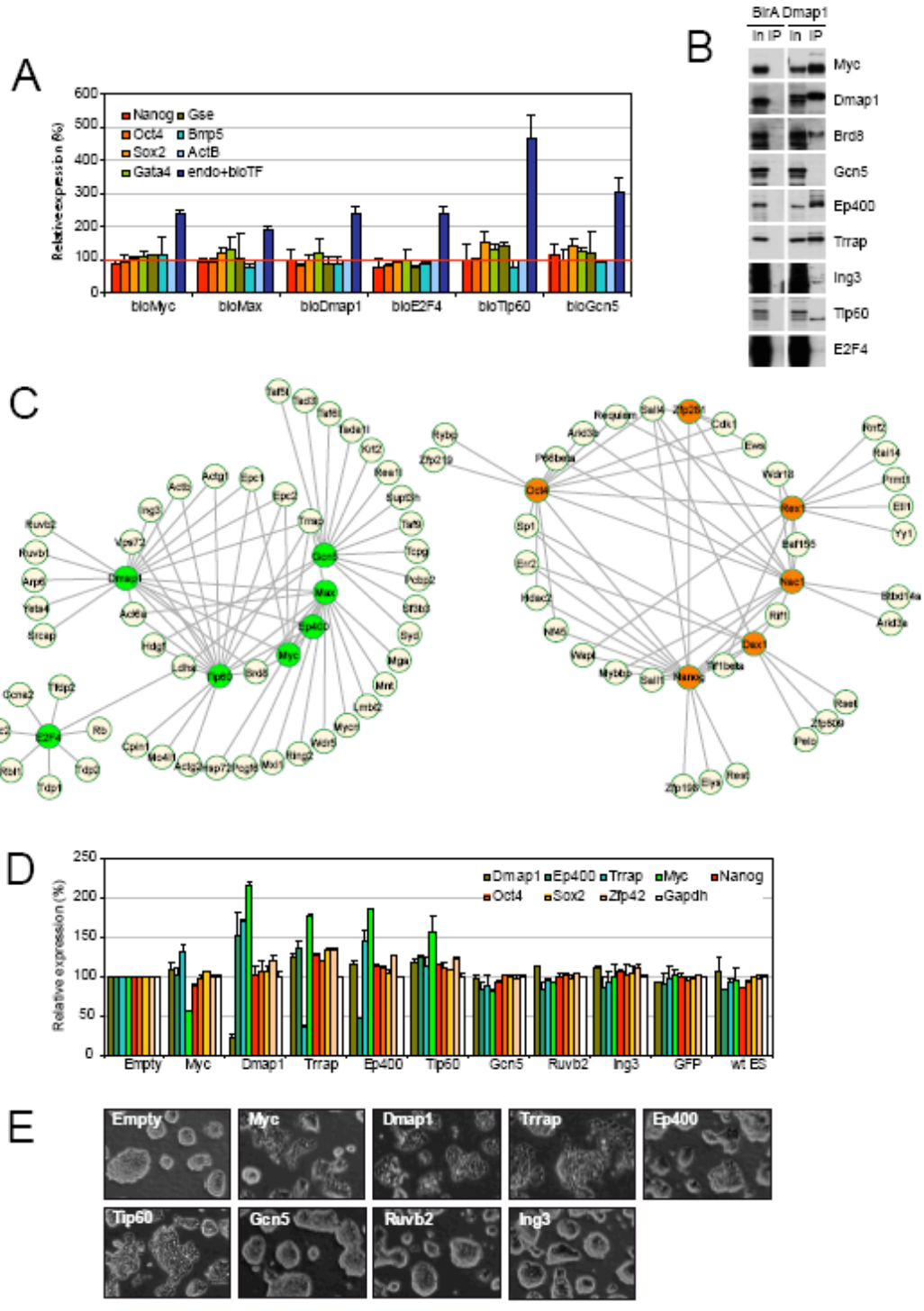


Figure S2

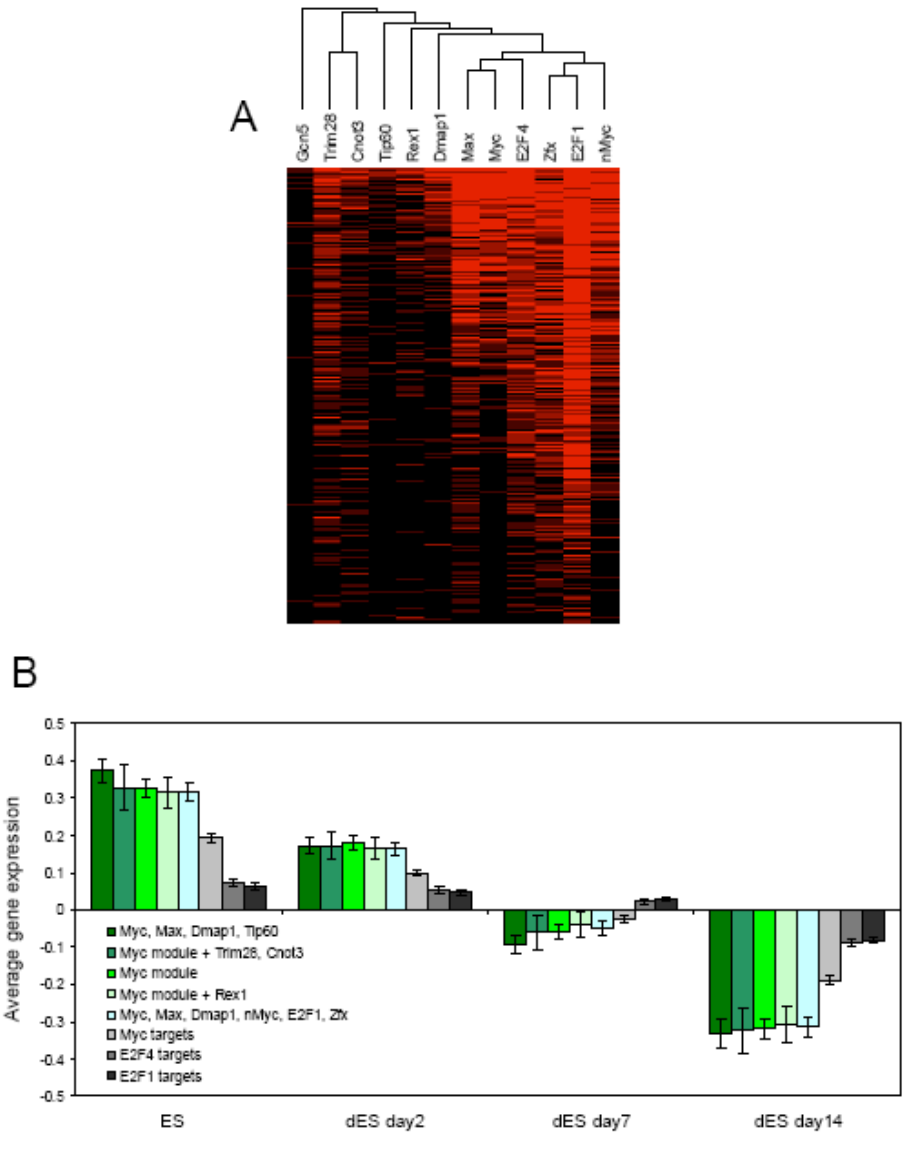

A
Gcn5
Trim28
Cnot3
Tip60
Rex1
Dmap1
Max
Myc
E2F4
Zfx
E2F1
nMyc
B
Average gene expression
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
Myc, Max, Dmap1, Tip60
Myc module + Trim28, Cnot3
Myc module
Myc module + Rex1
Myc, Max, Dmap1, nMyc, E2F1, Zfx
Myc targets
E2F4 targets
E2F1 targets
ES
dES day2
dES day7
dES day14

Figure S3

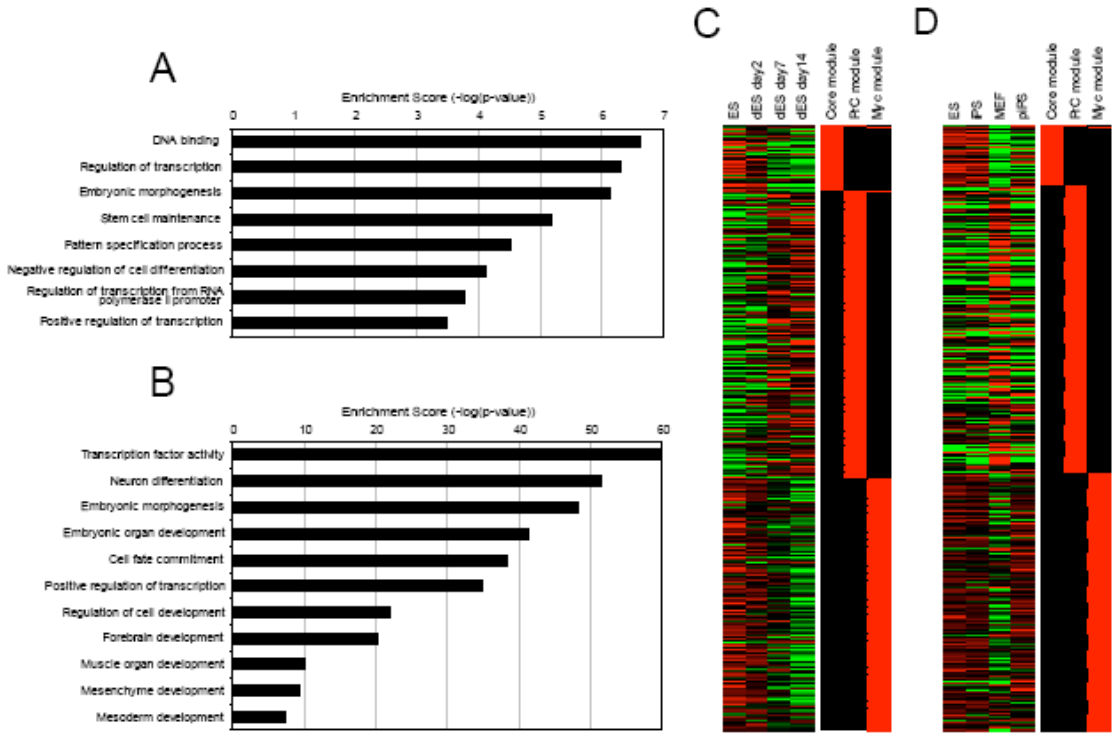
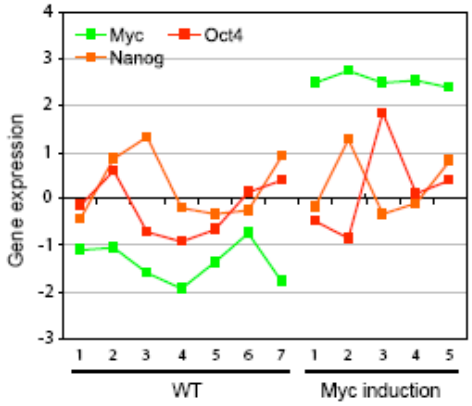

A
Enrichment Score (-log(p-value))
0
1
2
3
4
5
6
7
DNA binding
Regulation of transcription
Embryonic morphogenesis
Stem cell maintenance
Pattern specification process
Negative regulation of cell differentiation
Regulation of transcription from RNA polymerase II promoter
Positive regulation of transcription
B
Enrichment Score (-log(p-value))
0
10
20
30
40
50
60
Transcription factor activity
Neuron differentiation
Embryonic morphogenesis
Embryonic organ development
Cell fate commitment
Positive regulation of transcription
Regulation of cell development
Forebrain development
Muscle organ development
Mesenchyme development
Mesoderm development
C
ES
dES day2
dES day7
dES day14
Core module
PC module
Myc module
D
ES
iPS
MEF
piPS
Core module
PC module
Myc module

Figure S4

A


Myc
Oct4
Nanog
Gene expression
4
3
2
1
0
-1
-2
-3
1 2 3 4 5 6 7
1 2 3 4 5
WT
Myc induction

B

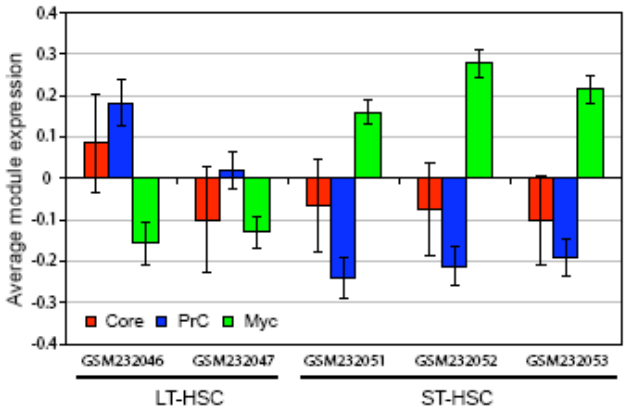
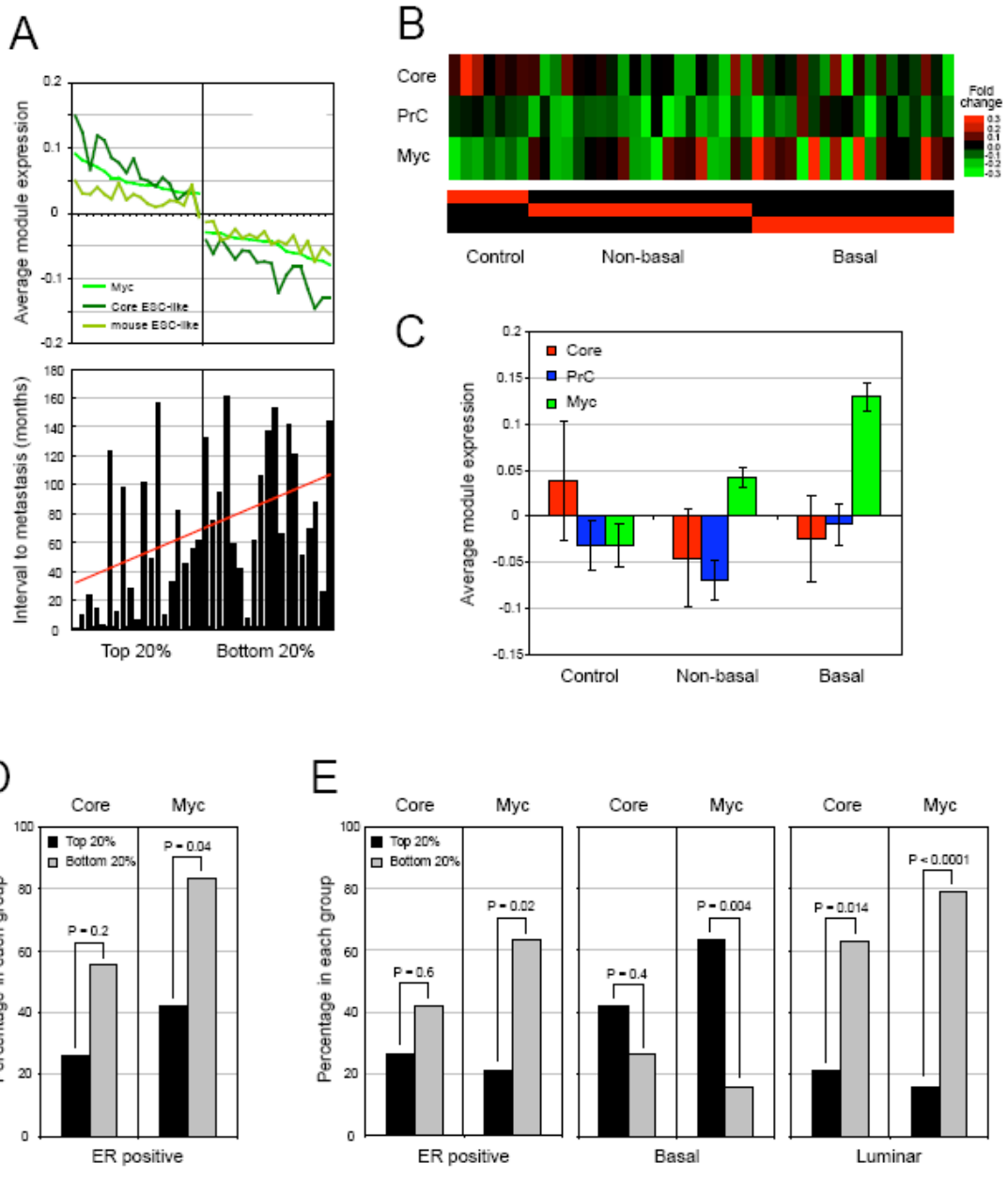

Average module expression
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Core
PrC
Myc
GSM232046
GSM232047
GSM232051
GSM232052
GSM232053
LT-HSC
ST-HSC

Figure S5


A
Average module expression
Myc
Core ESC-like
mouse ESC-like
Interval to metastasis (months)
Top 20%
Bottom 20%
B
Core
PrC
Myc
Fold change
Control
Non-basal
Basal
C
Core
PrC
Myc
Average module expression
Control
Non-basal
Basal
D
Core
Myc
Top 20%
Bottom 20%
P = 0.2
P = 0.04
Percentage in each group
ER positive
E
Core
Myc
Core
Myc
Core
Myc
Top 20%
Bottom 20%
P = 0.6
P = 0.02
P = 0.4
P = 0.004
P = 0.014
P < 0.0001
Percentage in each group
ER positive
Basal
Luminar

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Myc-centered protein-protein interaction network, Related to **Figure 1**

(A) Expression analysis of pluripotency and differentiation marker genes is performed using cell lines expressing biotin-tagged proteins. Relative gene expression levels compared to the levels in BirA expressing ES cells are shown. Data are represented as mean ± SEM.

(B) Input nuclear extract is treated with DNaseI/RNaseA for 1 hour, and the intercalating agent ethidium bromide is added to eliminate indirect protein-protein interaction mediated by nucleic acids. Strong protein-protein interactions of Dmap1 and other factors shown in **Figure 1C** are essentially recreated by co-immunoprecipitation.

(C) Depiction of the features of the Myc-centered protein-protein interaction network (left panel) and the previously identified Nanog-centered protein-protein interaction network (right panel) in ES cells. Proteins with green and red circles represent the biotin-tagged proteins used in affinity purifications.

(D-E) shRNA-based knockdown of Myc, Gcn5, and NuA4 complex proteins including Dmap1, Trrap, Ep400, Tip60, Ruvb2, and Ing3 is performed in ES cells. Relative expression of ES cell marker genes (Oct4, Sox2, Nanog, and Zfp42) and Myc/NuA4 complex genes (Dmap1, Ep400, Trrap, and Myc) upon knockdown of each factor is shown (D). Data are represented as mean ± SEM. ES cell morphology upon knockdown of each factor is shown (E). Empty (empty vector), GFP, and wtES (untreated) serve as controls.

Figure S2. Myc cluster and gene sets co-bound by factors in Myc cluster, Related to **Figure 2** and **Figure 4.**

(A) Hierarchical clustering of factors in the Myc cluster shown in **Figure 2** based on their target co-occupancy. Total 12154 genes which are targets of at least one of the 12 factors in the Myc cluster are tested.

(B) Targets co-occupied by different combinations of factors in the Myc cluster (**Figure 2**) are identified (listed in **Table S3**), and average gene expression values (log_2) of gene sets upon ES cell differentiation are plotted using the data set used in **Figure 4D**. Tested gene sets are 1) Myc, Max, Dmap1, and Tip60 common targets, 2) Myc module + Trim28, Cnot3 (Myc, Max, nMyc, Dmap1, E2F1, E2F4, Zfx, Trim28, and Cnot3 common targets), 3) Myc module (Myc, Max, nMyc, Dmap1, E2F1, E2F4, and Zfx common targets), 4) 7 factors in Myc module and Rex1 common targets, 5) Myc, Max,

Dmap1, nMyc, E2F1, and Zfx common targets, 6) all Myc targets, 7) all E2F4 targets, and 8) all E2F1 targets. Data are represented as mean ± SEM.

Figure S3. ES cell modules, Related to **Figure 4** and **Figure 5**
(A-B) Enrichment of pathways and Gene Ontology (GO) is tested using two groups of genes; genes in the Core module (A), and the PrC module (B).
(C-D) Averaged gene expression values (log_2) of ES cell modules upon ES cell differentiation shown in **Figure 4D** (C) as well as in various cell types, such as ES cells (ES), iPS cells (iPS), MEFs (MEF), and partial iPS cells (piPS) shown in **Figure 5A** (D) are re-plotted using individual gene activities as heat map graphs. Total 772 genes (C) and 1004 genes (D) are shown.

Figure S4. Nanog, and Oct4 expression levels upon Myc induction in human epithelial cells, and ES cell module activities in hematopoietic stem cells, Related to **Figure 5B** and **Figure 6**
(A) mRNA expression levels (log_2) of two individual ES cell core factors, Nanog and Oct4 are shown using the same expression data set used in **Figure 5B**. Individual array experiments (depicted as numbers) are shown. Level of Myc mRNA is also shown as a control.
(B) Average gene expression values (log_2) of ES cell modules (Core, Prc, and Myc) are tested in long-term hematopoietic stem cells (LT-HSC: Lin-cKit+Sca1+CD34-CD135-), and short-term hematopoietic stem cells (ST-HSC: Lin-cKit+Sca1+CD34+CD135-) using a previously published expression array data set (Ficara et al., 2008). Individual arrays are used for the analysis, and data are represented as mean ± SEM.

Figure S5. ES cell Core and Myc module activities in human breast cancer subtypes, Related to **Figure 7**
(A) Using the data set shown in Figure 7D, average gene expression values (log_2) of the previously defined Core ESC-like gene module (Core ESC-like), and mouse ESC-like gene module (mouse ESC-like) are calculated and plotted with the Myc module activity (Myc). Black bars (bottom panel) represent the corresponding interval to metastasis (months; same as shown in **Figure 7D** bottom). (B-C) Average gene expression values (log_2) of ES cell modules (Core, Prc and Myc) are tested using previously published human breast cancer subtype data including basal-like cancers (Basal), non-basal like cancers (Non-basal), and control samples (Richardson et al., 2006). Module activities of each patient sample (total 45 samples) are shown (B). Averaged module activities

within the sample groups (Control, Non-basal, and Basal) are shown as mean ± SEM (C).
(D) For each of the tested groups shown in **Figure 7C**, and **7D** (total 4 groups; samples with the top 20% and bottom 20% of Core module activity, and samples with the top 20% and bottom 20% of Myc module activity), percentage of estrogen receptor (ER) positive samples in each group are shown. P-values are calculated using Student's T-tests.
(E) Using an independently published data set containing 96 human breast cancer patients (Hu et al., 2006), Core and Myc module activities (average gene expression values (log_2)) are calculated. Four groups of samples (Core module activity: top and bottom 20% each; Myc module activity: top and bottom 20%, each; 19 samples in each group) are further analyzed. Percentage of corresponding ER positive samples (left panel), basal-like tumor samples (middle panel), and luminal tumor samples (right panel) within each group is plotted. P-values derived using Student's T-tests.

SUPPLEMENTAL TABLES

Supplied as Exel files

Table S1. Summary of biotin tagged proteins and their interacting partner proteins, Related to **Figure 1**.
Table S2. Summary of chromosomal target genes of transcription factors, and histone modification signatures, Related to **Figure 2**.
Table S3. Summary of ES cell modules with their human orthologues, and additional gene sets tested, Related to **Figure 4**, **5**, **6**, and **7**.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

ES Cell Lines and Culture

Mouse J1 ES cell lines were maintained in ES medium. DMEM (Dulbecco's modified Eagle's medium) is supplemented with 15% fetal calf serum, 0.1 mM β-mercaptoethanol, 2 mM L-glutamine, 0.1 mM nonessential amino acid, 1% of nucleoside mix (Sigma), 1000 U/ml recombinant leukemia inhibitory factor (LIF; Chemicon), and 50 U/ml Penicillin/Streptomycin. Expression of each biotin-tagged

transcription factor was confirmed by western blotting with anti-streptavidin–HRP or native antibody (see below) against the protein of interest.

Protein complex pull down and mass-spectrometry

One-step affinity purification with streptavidin-agarose was performed. In brief, the harvested cells are allowed to swell in a hypotonic solution containing 10 mM HEPES, 1.5 mM MgCl2, 10 mM KCl, 1 mM DTT, 1 mM PMSF and protease inhibitors. The cytoplasmic fraction is shed by making 4 passes in 26 gauge needle and syringe. For mass-spectrometry, 15-20 mg of nuclear protein from ES cells expressing BirA only (control) or both BirA and the biotin-tagged protein of interest (sample) were first pre-cleared with 150-200 µl of protein-G agarose beads for 1 hour followed by incubation in equivalent volume of streptavidin agarose beads overnight rocking at 4°C. Nuclear protein concentration is adjusted to 2 mg/ml in Tris-buffered saline (TBS) with the final 350 mM NaCl and 0.3% Nonidet P40 during the purification. The precipitated protein-beads were washed 4 times each for 15 minutes, eluted by boiling in Laemmli buffer for 5 min, and resolved on a 10% SDS-polyacrylamide gel. Whole lane LC-MS/MS sequencing and peptide identification were performed at the Taplin Biological Mass Spectrometry Facility at Harvard Medical School. At least two biological replicates were performed for each factor. Definition of high confidence interaction proteins within the complex is as follows; Proteins predicted by at least 3 peptides sequences from two independent purifications were considered. Proteins identified from the sample predicted by at least three times more peptide sequences than the proteins predicted from BirA expressing control cells were also considered as high confidence complex proteins. Tested biotin-tagged proteins and their high confidence interacting partner proteins with number of peptides are summarized in **Table S1**.

Protein complex pull down and Western blot

For immunoprecipitation followed by Western blotting assays shown in **Figure 1C**, the pulldown eluate from 1 mg of nuclear protein and 2% of the input were loaded in each lane. To verify that the resulting protein-protein interaction is not contaminated with indirect nucleic acid interactions, we used method described previously (Woo et al., 2008) where the input is pre-treated with DNase I (1 µg/ml), RNase A (1 µg/ml) and ethidium bromide (50 mg/ml) (**Figure S1B)**. The antibody used for Western blots are as follows; anti-Myc (Cell Signaling, #9402), anti-Trapp (Cell Signaling, #P2032), anti-Brd8 (Abcam, ab17958), anti-Ep400 (Abcam, ab70301), anti-Ing3 (Abcam, ab57388), anti-Dmap1 (Santa Cruz, SC-47583), anti-Gcn5 (Santa Cruz, SC-20698), anti-Max

(Cell Signaling, #4732), anti-Tip60 (Santa Cruz, SC-5725), E2F4 (Santa Cruz, SC-1082).

Chromatin Immunoprecipitation (ChIP)

For bioChIP reactions, streptavidin beads (Dynabeads MyOne Streptavidin T1) were used for the precipitation of chromatin, and 2% SDS was applied for one of the washing steps. At least three biological replicates were performed, and bioChIP reactions from BirA expressing J1 ES cells were used as a reference group. Conventional ChIP reactions were performed as described previously (Kim et al., 2005) with 1:100 (v/v) dilution of following antibodies; anti-acetyl-Histone H3 (upstate, #06-599), and anti-acetyl-Histone H4 (upstate, #06-866). Input genomic DNA was used for the reference sample.

Microarray and Data Processing

For ChIP-chip, ligation-mediated PCR was performed to amplify ChIP samples as described previously (Ren et al., 2000) and microarray hybridizations were performed on the Affymetrix GeneChip Mouse promoter 1.0R arrays as described in Affymetrix protocol. Microarray hybridization, incubation, washing, and scanning were performed at the Microarray Core Facility at the Dana Farber Cancer Institute. Model-based Analysis of Tiling-array (MAT) with p-value = 1.00E-6 was applied to predict the target loci (Johnson et al., 2006). Genomic regions between 8 kb upstream and 2 kb downstream of transcription start site (TSS) of well annotated genes from mouse genome annotation released in March 2006 (mm8) was used to assign the target genes. Raw and process ChIP-chip data set was deposited on the public server, Gene Expression Omnibus (GEO), under the accession number of GSE20551.

ChIP-chip or ChIP-sequencing data sets used for the analyses shown in **Figure 2** were obtained from the following sources;

1. http://www.nature.com/nature/journal/v441/n7091/suppinfo/nature04733.html (Boyer et al., 2006)*
2. GSE11431 (Chen et al., 2008)*
3. GSE11329 (Kim et al., 2008)*
4. GSE11283 (Hu et al., 2009)*
5. GSE15388 (Shen et al., 2008)*
6. GSE14654 (Ding et al., 2009)*

Gene expression profiling data sets used for the analyses were obtained from the following sources;

Figure 4: GSE3231(Boyer et al., 2006)*

Figure 5A: GSE14012 (Sridharan et al., 2009)*

Figure 5B: GSE3151 (Bild et al., 2006)*

Figure 6A: GSE13690 (Somervaille et al., 2009)*

Figure 6B: GSE13692 (Somervaille et al., 2009)*

Figure 7A, and **7B**:
http://jco.ascopubs.org/cgi/content/full/JCO.2005.03.2375/DC1
(Sanchez-Carbayo et al., 2006)*

Figure 7C, 7D, and **7E**; **Figure S5A**, and **S5D**:
http://www.rii.com/publications/2002/vantveer.html
(van 't Veer et al., 2002)*

Figure S4B: GSE9189 (Ficara et al., 2008)

Figure S5B, and **S5C**: GSE3744 (Richardson et al., 2006)

Figure S5E: https://genome.unc.edu/ (Hu et al., 2006)

'*' References are listed in the main text

Each expression microarray data set was generated by independent laboratories using various microarray platforms. To rule out the possible bias due to the sample preparation and different microarray platforms, normalization and data analysis were performed within each data set independently. Briefly, for each expression data set, individual probe intensity of each array was divided by the averaged probe intensity across all arrays within the data set, then each value was log (base 2) transformed. For normalization, first, average expression value of all probes in each array was calculated. Then for each array, expression value of each probe was subtracted by the averaged expression value. By doing so, average expression value of all probes in each array in each expression data set will be zero.

For the clustering, and Treeview image shown in **Figure 2**, we first performed un-supervised hierarchical clustering across the transcription factors based on their target co-occupancy using Cluster software (Eisen et al., 1998). For the target correlation map shown in **Figure 2D**, correlation coefficients for each pair of factors were calculated based on their target co-occupancy. For the target distance map shown in **Figure 2E**, in-house Perl script was used.

SUPPLEMENTAL REFERENECES

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.

Ficara, F., Murphy, M.J., Lin, M., and Cleary, M.L. (2008). Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. Cell Stem Cell *2*, 484-496.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L.*, et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci U S A *103*, 12457-12462.

Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nat Methods *2*, 47-53.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E.*, et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121-132.

Woo, A.J., Moran, T.B., Schindler, Y.L., Choe, S.K., Langer, N.B., Sullivan, M.R., Fujiwara, Y., Paw, B.H., and Cantor, A.B. (2008). Identification of ZBP-89 as a novel GATA-1-associated transcription factor involved in megakaryocytic and erythroid development. Mol Cell Biol *28*, 2675-2689.